PATHWAYS TO THE SLV COLLECTION

Working Paper 8

Vivienne Waller
September 2009

The Searchers

Institute for Social Research
Swinburne University of Technology
Acknowledgements

This Working Paper is part of the Australian Research Council project The Searchers. We would like to thank the many staff at the State Library of Victoria for their assistance and suggestions, in particular Sue Hamilton, Ian Patterson, Janice van de Velde, Anne Beaumont and Kirstie McRobert. This research was funded by Australian Research Council grant LP077215, conducted under the guidance of Chief Investigators Professor Julian Thomas and Professor Denise Meredyth and with the assistance of Dr Ian McShane. The Institute for Social Research takes all responsibility for the content. The views of the authors are not necessarily those of the State Library of Victoria.
CONTENTS

Executive Summary ........................................................................................................... 1
Background to the analysis ............................................................................................... 2
Arriving at the collection versus searching the collection ............................................. 2
Methodology...................................................................................................................... 3
Findings .............................................................................................................................. 5
Discussion .......................................................................................................................... 10
Concluding comments: the introduction of Primo ...................................................... 12
Executive Summary

This report compares accessing the website of the State Library of Victoria (SLV) via Internet search results with accessing its collection via the SLV catalogue. It compares estimates of the numbers of people who access the collection via each pathway. It also compares the subject of the query that leads to the website with the subject of the queries that members of the public type into the library catalogue. Although there are a range of caveats to the analyses in this report, the following findings are clear:

- In May 2008, the catalogue was still the main pathway to the SLV collection. It was particularly important for users wanting to find out about contemporary issues, cultural practice or art, or wanting to look up particular books or authors.

- Internet searches on history and places provide a significant pathway to the SLV’s collection. SLV collection material relating to history or places is more likely to be arrived at via search engine results than through catalogue queries.

- It appears that search engines are opening up the SLV collection, particularly Victoria-specific material, to people who would not choose to use the SLV catalogue.
Background to the analysis

The Searchers project examines the strategic challenges for major public libraries presented by the online information environment. The project has three components: research on current information-seeking and the role of public libraries in online information provision, analysis of wider developments in online environments and search models, and evaluation of the policy and strategic implications for public libraries highlighted by the research. The project will provide guidance for the State Library of Victoria (SLV or State Library) in developing appropriate models of information provision and contribute new thinking on the role and position of major libraries in the digital age.

This paper contributes to the series of Working Papers produced as part of the Searchers project. Other titles in the series are:

- Searchers Working Paper 3: Virtual Visitors to the website of the State Library of Victoria, August 2008
- Searchers Working Paper 4: Accessing the State Library of Victoria through iTunes University (beyond the campus), December 2008
- Searchers Working Paper 6: Online visits to large public libraries in Australia - a comparison, December 2008

This report contributes to the research on current information-seeking, analysing the pathways to the SLV collection, comparing the search engine queries that brought visitors to the SLV website (from Searchers Working Paper 3: Virtual Visitors to the website of the State Library of Victoria) with the content of catalogue search terms analysed in Searchers Working Paper 7: What do people search for on the State Library catalogue?

Arriving at the collection versus searching the collection

In the past, the only way to access a library’s collection was to browse the physical collection, search or browse the library catalogue or ask a librarian. The Internet makes possible new ways of accessing the library collection. Through clicking on a link in a webpage, people may arrive at an item in the library collection without even consciously choosing the library as their destination. Furthermore, with many libraries making their collection metadata progressively available to Internet search engines, links to a library catalogue or collection can now appear in search engine results. Figure 1 depicts these alternative ways of accessing the collection of the State Library of Victoria.
This report compares accessing the State Library of Victoria (SLV) via internet search results with accessing its collection via the SLV catalogue. It compares estimates of the numbers of people who access the collection via each pathway. It also compares the subject of the query that leads to the website with the subject of the query typed into the library catalogue by the public. In doing this, it draws together the results of an analysis of what people search for on the SLV Main catalogue (Searchers Working Paper 7, 2009) with an analysis of the terms that brought people to the SLV website via search engine results (Searchers Working Paper 3, 2008).

**Methodology**

**Data sources**
The main catalogue of the State Library of Victoria contains records for books, magazines, newspapers, electronic books and journals, video recordings, music, maps and oral history as well as records of several thousand websites. The system used by the State Library at the time of analysis was a Voyager (ExLibris) catalogue. By 2008 this catalogue had more than 1.3 million items. The content analysis in Working
Paper 7 is based on a random stratified sample of 1,000 transaction logs from the main catalogue for the month of May 2008. The sample was restricted to members of the public using the catalogue (in other words, it excluded SLV staff searches). The sample was also restricted to the first search undertaken in a session. This reduced double counting of what is essentially part of the one search. This also meant that the sample only included search queries typed in by the user rather than those that were the result of clicking in the resulting record after an initial typed in query. Lastly, the content analysis of the search queries was restricted to those search types that were considered to be about a general topic rather than a specific item and the sample was stratified according to the prevalence of each included search type. (For a fuller description of the methodology, refer to Searchers Working Paper 7.)

While title searches and author browse search types were excluded from the content analysis in Searchers Working Paper 7, they have been added back into the results reported in this paper as queries relating to books or authors.

The data in this paper on search terms that resulted in a visit to the SLV website is sourced from Hitwise data on Australian internet use. (For a description of the Hitwise methodology, refer to Searchers Working Paper 3.)

**Subject classification**

Each of the search queries in the sample of catalogue queries and the sample of web search terms was manually examined and coded. These codes were created to most closely describe the content of the search queries. The data were combed over numerous times, sorted by code and recoded as necessary. Unless the meaning was immediately obvious, each search query was looked up in Google and/or the SLV catalogue. In this way, each search query received one or more tags which were progressively refined to indicate the subject of the search.

In Searchers Working Paper 3, there was a substantial group of search terms (9%) that were recorded as leading to the SLV website, but seemed very unlikely to have resulted in search results that linked to the SLV website. These ‘unrelated’ queries were excluded from the following analysis as it is likely that these visitors to the SLV website used tabbed browsing or a bookmark to go directly to the SLV website. In addition, the analysis of unique search queries that brought people to the SLV website included 12% tagged as ‘reference’. For this paper, search terms with this tag were recoded to the subject of the reference query.

The coding inevitably involves some subjective decisions about the appropriate category in which to place a particular term. Although the contribution of each query to the whole is small, there is a margin of imprecision to the classifications. In interpreting the data, one should focus on the overall pattern, rather than the precise size of each category.
**Findings**

The results of combining the two analyses are shown in Figure 2. This compares the unique terms typed into the catalogue (the first search of a session) with the terms that brought people to the SLV website.

![Figure 2: Comparison of catalogue queries with search queries that led to SLV website – May 2008](image)

For the period under investigation, only SLV collection items that included digitised images were visible to search engines\(^1\). This included manuscripts, maps, pamphlets, artworks and photographs. Catalogue records for books were theoretically ‘visible’ to search engines via the Libraries Australia website, but could not be accessed directly from search engine results.

It is not possible to distinguish between those search engine queries that resulted in links to the SLV collection and those that resulted in links to the SLV website. During the period under study (May 2008), as well as containing material about events and exhibitions at the SLV, the SLV website contained a range of other material (for example material relating to the curriculum of Senior secondary students (Ergo), information guides on conservation, and information pages on explorers Burke and Wills and artist Eric Thake). The SLV website also hosts the Latrobe Journal which contains articles which use the Library's collections as source material. Hence most of the material on the SLV website relates to the SLV collection in some way. It is fair to assume that queries relating to SLV and other libraries and some of

\(^1\) This was through the addition of Dublin core metadata.
the queries relating to books and authors would have resulted in links to areas of the SLV website that did not relate directly to the collection.

It should also be noted that aggregating the queries and search terms into subjects hides important differences between the two data sources. For example Figure 2 shows that two fifths of catalogue searches were related to books, authors or newspapers. The library catalogue has the facility for searching book titles or authors which means that the book-related searches would have been mostly for book titles. In contrast, searching on the Internet for a book title would only have lead indirectly to the record of that book in the SLV collection via Libraries Australia and only 0.07% of all visits to the SLV website were via this route (Source: Hitwise, May 2008). The book-related web search queries that led to the SLV site were about books rather than book titles.

Despite the caveats, Figure 2 gives a rough indication of the subject of collection material being reached via search engines and the subject of material being accessed via the catalogue.

- Those who arrived at the SLV website via a search engine were much more likely to be searching for material relating to history and places than those who searched the catalogue.
  - Almost one quarter of visitors to the SLV website who came via a search engine had a query relating to a place/building compared to just over 5% of catalogue users whose query related to a place/building.
  - More than 15% of visitors to the SLV website who came via a search engine had a query relating to history compared to just over 5% of catalogue users who had a query relating to history.

- Those who searched the catalogue were more likely to be looking for material relating to contemporary issues, and art/music than those who arrived at the SLV website via a search engine.

These observations about the *topics* raise the question of whether this means that *people* are more likely to access particular subject areas of the collection via search results or the catalogue. This, of course, depends on the numbers of people accessing the SLV web via search results and the numbers of people using the SLV catalogue.

Figure 3 shows the number of visits to the SLV website (via a search engine) and searches on the main catalogue for the period 2006-2008.
In May 2008, 106,434 people clicked on the SLV website in search engine results (Source: Webtrends). In the same month, members of the public typed 117,507 search queries into the catalogue. The graph indicates that, between January 2006 and May 2008, the number of searches on the main catalogue was of the same order as the number of web searches that led to SLV.

However, the following factors need to be borne in mind when interpreting Figure 3.

- The search queries that led an internet user to the SLV website included queries that related to the State Library of Victoria itself rather than its collection.
- The number of catalogue searches does not represent the number of times a user researched a subject on the catalogue. Many users of the catalogue have to conduct several queries on the one topic in order to obtain the most relevant results (see Working Paper 5). These modifications of the initial query are really part of the one search.
- The number of main catalogue searches does not include the searches on the pictures, manuscript and Victorian Government publications catalogue.

Figure 4 addresses some of these issues to enable a more direct comparison for May 2008.
Figure 4: Comparison of numbers of visitors to slv web via search engines and numbers of catalogue searches: May 2008

In Figure 4:

- The searches that led to the SLV collection are assumed to be the web searches that led to the SLV excluding queries about the SLV or other libraries and queries coded as ‘unrelated’.

- The number of searches on all the catalogues of the SLV are included; main, pictures, manuscript and Victorian Government Publications. The number of sessions is also depicted. While the number of searches is an overestimate of the number of times a user researched a subject on the catalogue, the number of sessions is an underestimate. There were catalogue users who typed in more than one query, not because of any difficulty in retrieving relevant results, but because they were researching more than one topic or looking for more than one specific item. (See Searchers Working Paper 5). It is not possible to know how many were in each group. Hence the number of catalogue queries for comparison with the search terms that led to the SLV website lies somewhere between the number of catalogue searches and the number of catalogue sessions.
The main point illustrated in Figure 4 is that in May 2008, it appears that the catalogue was still the main pathway to the SLV collection.

While Figure 2 compared the subject of catalogue queries with the subject of web queries that brought people to the SLV website, Figure 5 indicates whether people are more likely to access particular subject areas of the collection via internet search or the catalogue. This graph applies the proportions depicted in Figure 2 to the numbers in Figure 4 (main catalogue only). As with Figure 4, the number of catalogue queries for comparison with the search terms that led to the SLV website lies somewhere between the number of catalogue searches and the number of catalogue sessions.

Figure 5 indicates that certain topics were more likely to be searched in the main catalogue than appear as links to the SLV website in search engine results. These topics were contemporary issues, cultural practice, art/music, business and health. This shows the importance of the catalogue as a means of accessing the collection, particularly for users wanting to find out about these topics or looking up particular books or authors.

Conversely, Figure 5 also indicates that particular topics are more likely to appear as
links to SLV in search engine results than they are to be searched in the main
catalogue\textsuperscript{2}. In Searchers Working Paper 7, it was reported that catalogue queries about
history were less likely to be about Victorian history and more likely to be about the
history of another country. This was explained in terms of the Australiana index and
the La Trobe Reading room as main points of entry to the SLV’s collection on
Victorian history. Figure 5 shows that web search queries are a significant point of
entry to the SLV’s collection on Victorian history; as described in Searchers Working
Paper 3, most history queries related specifically to Victoria. Figure 5 also shows that
web search queries are a significant point of entry to the SLV’s collection on places;
as described in Searchers Working Paper 3, queries on places or particular buildings
were almost exclusively related to Victorian places or buildings.

\textit{Discussion}

This report provides empirical evidence that, as would be expected, visitors arriving
via a search engine and catalogue users are accessing different parts of the collection.
A visitor via a search engine is much more likely to access information on a place or
history, whereas a catalogue user is much more likely to access material relating to
contemporary issues, art\textsuperscript{3} and cultural practice.

The reasons for these differences relate to people’s preferences regarding information
sources, the visibility of the collection to search engines and to the workings of the
search engine algorithm\textsuperscript{4}. One can type anything as a search query into a catalogue. 
However, there are several conditions that need to be satisfied in order for someone to
arrive at the SLV website via search engine results. Firstly, the website needs to be
visible to search engines and as mentioned, at the time of research, only those items
associated with digitised images were visible to search engines. Secondly, a link to
the SLV website has to appear in the first page or so of the search engine results\textsuperscript{5}. 
The level of visibility of the SLV website in search results is partly a reflection of the
content of the SLV website (accessible to search engines), and partly a reflection of
what else is on the web on that subject. For example, if one is using a search engine

\textsuperscript{2} Figure 5 excludes searches on the pictures catalogue as these were not coded by subject. It is likely
that these searches would have been mainly related to places, buildings, artists and history, but the
numbers would have been insufficient to change the overall patterns depicted in Figure 5. As Figure 4
shows, there were just over 10,000 sessions conducted in the pictures catalogue in May 2008.

\textsuperscript{3} More than 90\% of the queries labelled art/music were related to art (including architecture).

\textsuperscript{4} According to Hitwise, 80\% of people using a search engine in Australia in May 2008 were using
Google.

\textsuperscript{5} Almost two–thirds (62 percent) of search engine users do not go past the first page of results and less
than 10 percent go past the first three pages of results (“iProspect Search Engine User Behaviour Study
4 June 2008.)
to find information on a contemporary issue, it is likely that a range of websites devoted to that particular topic, including the relevant Wikipedia entry, will dominate the results. Lastly, the user has to choose to click on the link to the SLV website. We do not have data on how many times a user was presented with links to the SLV in the search results, only on how many times these links were followed. The data supports, although it does not prove, the hypothesis that internet users are more likely to choose a link to the State Library of Victoria collection when the subject of their search relates specifically to Victoria. The link in the search results does not identify the State Library of Victoria other than through the acronym ‘slv’. However, the web address visible in the search results does include “vic.gov.au” which is widely recognisable as associating the website with the government of Victoria and could be supposed to lend credibility to the result.

Another point regarding people’s preferences concerning information sources is that it is clear from the analysis of catalogue queries that the SLV collection is still deliberately chosen as a source of information. More than half of all main catalogue searches are for specific items (Source: Searchers Working Paper 7) while the most searched topics are contemporary issues, art and cultural practices. This suggests that these searchers are finding greater value in searching for these topics on the main catalogue than on the Internet and want to find books on these topics rather than websites. At present, a library catalogue is a more efficient way of finding a book about a topic than a general search engine.

This report includes the following descriptive example to illustrate how search engines are opening up parts of the SLV collection to people who may not have chosen to use the SLV catalogue.

**Searching for the Jerilderie letter**

Excluding search terms that relate to aspects of the ‘State Library of Victoria’, including permutations of “State Library of Victoria, Melbourne”, the search term that consistently brings the most visitors to the SLV is ‘Jerilderie Letter’ or ‘the Jerilderie Letter’ (Hitwise data: August 2006, May 2007 and Google Analytics). The Jerilderie letter, dictated by Ned Kelly, is a significant part of Victoria’s history and has been part of the collection of the State Library of Victoria since the year 2000.

Few catalogue users search for the Jerilderie letter. During May 2008, queries containing the word ‘Jerilderie’ occurred a total of 22 times; four times in the pictures catalogue, 18 times in the manuscripts catalogue and not at all in the main catalogue.

Typing “Jerilderie letter” into Google, brings up a link to the SLV digitised image of the first page of the Jerilderie letter ([http://www.slv.vic.gov.au/collections/treasures/jerilderieletter.html](http://www.slv.vic.gov.au/collections/treasures/jerilderieletter.html)). From there, a visitor can look at a digitised image of each page of the original letter as well as a transcript. This search query brings an average of 15 visitors per day to the SLV
collection. These visitors have not looked up Jerilderie letter on the SLV catalogue and it is quite possible that many of those who are discovering the SLV digitised version of the Jerilderie letter were not previously aware of the SLV collection.

**Concluding comments: the introduction of Primo**

As of September 2009, the SLV catalogue has been replaced with a next generation catalogue (Primo by Ex Libris). Primo turns the catalogue records into a format that is visible to search engines. This means that the entire catalogued collection of the SLV can potentially be arrived at via search engine results. If it is true that internet users are more likely to choose a link to the State Library of Victoria collection when the subject of their search relates specifically to Victoria, then it is possible that this will increase access to books about Victoria which are in the SLV collection. However, a user arriving at the collection in this way is only accessing the catalogue record. Unless these books are digitised, additional arrangements need to be made to access the content (the physical book). The tiny proportion of visits to the SLV website from Libraries Australia in May 2008, suggests that there is not likely to be a huge jump in access. However, an additional factor making the book collection visible is the May 2008 agreement between Google and the Library of Congress which enables Google books to link to Libraries Australia and hence to the SLV catalogue. It is likely, however, that both the catalogue and the internet will remain important pathways for accessing different parts of the SLV collection.

---